



University of Kentucky
UKnowledge

Theses and Dissertations--Public Health (M.P.H.
& Dr.P.H.)

College of Public Health

2018

Consultation Experience in Public Health

Xiangya Meng

University of Kentucky, xme224@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/cph_etds



Part of the [Public Health Commons](#)

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Meng, Xiangya, "Consultation Experience in Public Health" (2018). *Theses and Dissertations--Public Health (M.P.H. & Dr.P.H.)*. 217.

https://uknowledge.uky.edu/cph_etds/217

This Graduate Capstone Project is brought to you for free and open access by the College of Public Health at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Public Health (M.P.H. & Dr.P.H.) by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my capstone and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's capstone including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Xiangya Meng, Student

Dr. Heather Bush, Committee Chair

Dr. Sarah Wackerbarth, Director of Graduate Studies

Consultation Experience in Public Health

Xiangya Meng

MPH

Biostatistics

Contents

Project Description.....	2
My Role	3
Applicable SAS code.....	4
Format and Marco code.....	4
Analyzing and Reporting Code.....	8
Project Outcomes (Tables and Graphs)	12
Descriptive Tables and Graphs.....	12
Logistic Regression	20
Lessons Learned.....	25
Project Summary.....	26

Project Description

According to data from the National Youth Risk Behavior Survey (YRBS), rates of teen dating and sexual violence remains high, which causes severe problems for teenagers' health. In order to reduce the risk of sexual violence, some prevention programs are designed and conducted in past years. Green Dot program, one of the prevention programs, was developed to reduce dating and sexual violence, especially as students transition from high school into young adulthood. It is an active bystander-based randomized intervention trial in 13 pairs of high schools across the Bluegrass in Kentucky. However, many prevention programs were only followed for short-term effects. Extended follow up is necessary to detect changes in violence since a lasting behavior change needs long periods of time. Hence, life's Snapshot Project was designed, aiming to evaluate the long-term effects of Green Dot Intervention.

In a prospective longitudinal study, retention of participants over follow-ups would be challenging. In Life's Snapshot Project, over 11000 participants were recruited at baseline surveys for three years as 3 cohorts. Baseline surveys were collected in person at high school sites, receiving a high rate of response. Then each cohort was annually followed within the five-year study period. Cohort 1 was followed for five consecutive years, cohort 2 was followed for four consecutive years, and cohort 3 was followed up for three consecutive years. Follow-up surveys were collected via electronic approach and were only sent to people who responded in the first follow-up. However, the final survey was sent to everyone recruited at baseline and 1986 responses were obtained. Incentive structure was developed to increase response and retention rates. However, due to email access, internet access and dropping out, numerous participants were lost to follow-up.

This project aims to investigate the factors influencing response rates of follow-up by describing the demographics of people who only finished the baseline survey, people who came back for follow-up, and people who responded through follow-up duration. Another purpose is to investigate whether Green Dot program experience and college attendance make a difference in retention rates. Hence, cohort 1 was chosen as it has the completed five-year follow-ups.

Green Dot experience was divided into two groups, one group includes people who had experienced Green Dot Training, Green Dot Speeches, or heard about Green Dot program, while the other group is people who know nothing about this program. The intervention was also compared between different follow-up groups, even though the intervention and control groups were not clearly separated due to networking. Demographics analyzed include gender (male and female), residential area (urban, suburban, and rural), race (white or nonwhite), and poverty (divided by 15.8 and 17.8). Whether students took AP class was contained in the analysis as an indicator for college attendance.

My Role

This project was developed under Dr. Bush's supervision. The database is provided by Life's Snapshot Project.

My responsibilities:

1. Understanding the purposes and objectives of the project
2. Extracting subgroups of participants
3. Performing data cleansing, creating new variables or indicators
4. Merging multiple datasets
5. Performing data analysis, including descriptive analysis, Chi-square tests, and logistic regression by SAS, version 9.4, software (SAS Institute, Inc)
6. Coding format and macro to analyze data
7. Using macro to generate reports

Applicable SAS code

Format and Marco code

```
proc format;
value surind_a
1="Baseline Only"
2="At Least One Follow-up"
other='';

value surind_b
1="Baseline Only"
2="More Than Two Follow-ups"
other='';

value surind_c
1="Baseline Only"
2="Five Follow-ups"
other='';

value surind_d
1="At Least One Follow-up in Four"
2="Only Final"
other='';

value surind_e
1="Baseline only"
2="Final finished"
other=' ';

value gdmatn
1="Green Dot Training or Speeches"
0="Other Training or No Experience"
other='';

value gdmatyn
1="Green Dot Training, Speeches or Other"
0="No Experience"
other='';

value gender
1="Female"
2="Male"
other="";

value int
1="Intervention"
0="Control"
other='';

value resarea
1="Urban"
2="Suburban"
```

```

3="Rural"
other='';
value nonwhite
1="Non-white"
0="White"
other='';

value racegp
1="White"
2="Black"
3="Other Race"
other='';

value poverty
1="<=15.8"
2="15.8 - 17.8"
3=">17.8"
other='';

value apyn
1="Yes"
0="No"
other='';

/*
value col
1="At Least Plan to Attend 4-year College Once"
0="Never Plan to Go to College"
other='';
*/

value fupgp
0="Baseline only"
1="One follow-up completed"
2="Two follow-ups completed"
3="Three follow-ups completed"
4="Four follow-ups completed"
5="Five follow-ups completed";

value numfups
1="0"
2="1-2"
3="3-4";

run;

%macro dofreq(ds,sv,oneval,dvar,ivar,fmt,final);

/*Get Total Frequency*/
proc sort data=&ds out=tmp;
by &dvar;
run;

ods output OneWayFreqs=FRQ1;
proc freq data=tmp(where=(&ivar > .));
tables &dvar;
run;

```



```

data FRQ1;
length col $20;
set FRQ1;
col=compress(put(frequency,8.))||' ('||compress(put(percent,8.1))||"%");
if &dvar > .;
run;
/*transpose the result table, sort first*/
proc sort data=FRQ1;
by &dvar;
run;

proc transpose data=FRQ1 out=TFRQ1;
by &dvar;
var col;
run;

/*Get Total Column Frequency*/
ods output OneWayFreqs=FRQC;
proc freq data=tmp;
tables &ivar;
run;

ods output OneWayFreqs=FRQN;
proc freq data=tmp;
tables pat;
run;

data FRQC;
length colT parm $250;
set FRQN FRQC;
colT=put(frequency,8.);
if table="Table pat" then &ivar=-1;
parm="Total";
run;

proc transpose data=FRQC out=TFRQC(rename=(N1=COL1));
id &ivar;
var colT;
run;

/*Get General Association output*/
ods output chisq=CHISQ;

/*Get Frequency output*/
ods output crossTabFreqs=FRQS;
proc freq data=tmp(where=(&ivar > .));
tables &dvar*&ivar / chisq;
run;

data FRQS(where=(&dvar ne .));
set FRQS(where=(&ivar > .));
run;

data FRQS;
length col $20;
set FRQS;
*col=compress(put(frequency,8.))||' ('||compress(put(rowpercent,8.1))||"%");

```

```

col=compress(put(frequency,8.))||' ('||compress(put(colpercent,8.1))||"%)";
if &ivar > . and &dvar > .;
run;

proc sort data=FRQs;
by &dvar;
run;
proc transpose data=FRQs out=TFRQs;
by &dvar;
id &ivar;
var col;
run;

data CHI(keep=parm2 parm pvalue sv dfr);
length parm $250 pvalue $30 parm2 $10 sv 8.;
set chisq(where=(statistic='Chi-Square') keep=prob statistic df);
parm="&dvar";
parm2="&ivar";
sv=0;
if prob > . then do;
if &oneval=0 then pvalue=" ";
else pvalue=put(prob,pvalue6.);
dfr="DF="||trim(left(put(df,best.)));
end;
run;

/*merge tables with pvalue and frequency (percent)*/
proc sort data=TFRQ1;
by &dvar;
run;

proc sort data=TFRQs;
by &dvar;
run;

data TFRQs;
merge TFRQ1 TFRQs;
by &dvar;
run;

/*merge tables with pvalue and frequency (percent)*/
data TFRQs(drop=&dvar _NAME_);
length parm $250;
set CHI TFRQs(where=(&dvar>= 0));
if sv=0 then parm="&dvar";
else parm=put(&dvar,&fmt.);
run;

/*cumulate tables*/
data &final;
length sv 8. pvalue $30 ;
set &final TFRQS(in=a);
if a then do;
end;
run;

%mend;

```

Analyzing and Reporting Code

```
libname adata "S:\DATAQUEST\Requests\Snapshot\Data\ADATA";
option nofmterr;

%macro col(dsn,adata,ind);
data &dsn;
set &adata;
&ind=1;
run;

proc sort data=&dsn;
by st_id;
run;
%mend;

%col(c1b1,adata.ad_b1c1,basind);
%col(c1fu1,adata.ad_fulc1,fulind);
%col(c1fu2,adata.ad_fu2c1,fu2ind);
%col(c1fu3,adata.ad_fu3c1,fu3ind);
%col(c1fu4,adata.ad_fu4c1,fu4ind);
%col(c1final,adata.ad_c1final,finalind);

data merged;
merge c1b1 c1fu1 c1fu2 c1fu3 c1fu4 c1final;
by st_id;
ful_4=sum(of fulind,fu2ind,fu3ind,fu4ind);
fups=sum(of fulind fu2ind fu3ind fu4ind finalind);
*college=sum(of collsyn afths_5 afths_2_7 afths_3_7 afths_4_7
afths_f_7);/*plan to attend 4-year college after high school*/
*if college=0 then col=0; /*never plan to attend 4-year college*/
*else col=1; /*at least once plan to attend 4-year college*/
if GDMAT in (1,2) then GDMATN=1;/*know GREEN DOT*/
else GDMATN=0;/*know nothing about GREEN DOT*/
if GDMAT in (1,2,3) then GDMATYN=1;/*know GREEN DOT*/
else GDMATYN=0;/*know nothing about GREEN DOT*/
if fups=. then surind1=1;/*baseline only*/
else surind1=2; /*at least one follow-up*/
if fups=. then surind2=1;/*baseline only*/
else if fups>2 then surind2=2;/*more than 2 follow-ups*/
else surind2=.;
if fups=. then surind3=1;/*baseline only*/
else if fups=5 then surind3=2;/*five follow-ups*/
else surind3=.;
if ful_4>=1 then surind4=1;/*at least one follow-up in four*/
else if ful_4=. and finalind=1 then surind4=2;/*only final*/
else surind4=.;
if surind1=1 then surind5=1;/*baseline only*/
else if finalind=1 then surind5=2;/*finished final no matter what other
completed*/
else surind5=.;
pat=1;
if fups=. then fupgp=0;
else if fups=1 then fupgp=1;
else if fups=2 then fupgp=2;
```

```

else if fups=3 then fupgp=3;
else if fups=4 then fupgp=4;
else fupgp=5;
if ful_4=. and surind5=2 then numfups=1;
else if ful_4 in (1,2) and surind5=2 then numfups=2;
else if ful_4 in (3,4) and surind5=2 then numfups=3;
run;

%include "M:\MPH\2018 Fall Semester\CPH 608 Capstone\codes\Format and
Macro.sas";

ods output OneWayFreqs=FRQ;
proc freq data=merged;
table fupgp;
format fupgp fupgp.;
run;
data FRQ;
length col $20;
set FRQ;
col=compress(put(frequency,8.))||'| ('||compress(put(percent,8.1))||"%" );
run;

data final1 final2 final3 final4;
stop;

/*Baseline Only and At Least One Follow-up*/
%dofreq(merged,1.1,1,GDMATYN,surind1,gdmatyn.,final1);
%dofreq(merged,1.1,1,INT,surind1,int.,final1);
%dofreq(merged,1.1,1,GENDER,surind1,gender.,final1);
%dofreq(merged,1.1,1,RUCCGP,surind1,resarea.,final1);
%dofreq(merged,1.1,1,NONWHITE,surind1,nonwhite.,final1);
%dofreq(merged,1.1,1,POVGP,surind1,poverty.,final1);
%dofreq(merged,1.1,1,APYN,surind1,apyn.,final1);

/*Baseline Only and Five Follow-ups*/
%dofreq(merged,1.1,1,GDMATYN,surind3,gdmatyn.,final2);
%dofreq(merged,1.1,1,INT,surind3,int.,final2);
%dofreq(merged,1.1,1,GENDER,surind3,gender.,final2);
%dofreq(merged,1.1,1,RUCCGP,surind3,resarea.,final2);
%dofreq(merged,1.1,1,NONWHITE,surind3,nonwhite.,final2);
%dofreq(merged,1.1,1,POVGP,surind3,poverty.,final2);
%dofreq(merged,1.1,1,APYN,surind3,apyn.,final2);

/*Baseline Only and Final Finished*/
%dofreq(merged,1.1,1,GDMATYN,surind5,gdmatyn.,final3);
%dofreq(merged,1.1,1,INT,surind5,int.,final3);
%dofreq(merged,1.1,1,GENDER,surind5,gender.,final3);
%dofreq(merged,1.1,1,RUCCGP,surind5,resarea.,final3);
%dofreq(merged,1.1,1,NONWHITE,surind5,nonwhite.,final3);
%dofreq(merged,1.1,1,POVGP,surind5,poverty.,final3);
%dofreq(merged,1.1,1,APYN,surind5,apyn.,final3);

/*At Least One Follow-up in Four and Only Final*/
%final(final4);
%dofreq(merged,1.1,1,GDMATYN,surind4,gdmatyn.,final4);
%dofreq(merged,1.1,1,INT,surind4,int.,final4);
%dofreq(merged,1.1,1,GENDER,surind4,gender.,final4);

```

```

%dofreq(merged,1.1,1,RUCCGP,surind4,resarea.,final4);
%dofreq(merged,1.1,1,NONWHITE,surind4,nonwhite.,final4);
%dofreq(merged,1.1,1,POVGP,surind4,poverty.,final4);
%dofreq(merged,1.1,1,APYN,surind4,apyn.,final4);

proc freq data=merged;
table numfups;
format numfups numfups.;
run;

%global GP1 GP2 GP3 GP4 GP5 GP6;

proc sql;
select count (distinct st_id)
into:GP1
from merged(where=(surind1=1));/*Baseline Only*/

select count (distinct st_id)
into:GP2
from merged(where=(surind1=2));/*At Least One Follow-up*/

select count (distinct st_id)
into:GP3
from merged(where=(surind3=2));/*Five Follow-ups*/

select count (distinct st_id)
into:GP4
from merged(where=(surind5=2));/*Final Finished*/

select count (distinct st_id)
into:GP5
from merged(where=(surind4=1));/*At Least One Follow-up in Four*/

select count (distinct st_id)
into:GP6
from merged(where=(surind4=2));/*Only Final*/
quit;

%let GP1=%left(&GP1);
%let GP2=%left(&GP2);
%let GP3=%left(&GP3);
%let GP4=%left(&GP4);
%let GP5=%left(&GP5);
%let GP6=%left(&GP6);

%macro report(final,group1,group2,N1,N2,table);
proc report data=&final headline headskip nowindows split='|';
if &final=final1 %then %do;
columns sv parm _1 _2 pvalue;
%end;
%else %do;
columns sv parm _1 _2;
%end;
define sv / display "sv";
define parm / display " ";
define _1 / display "&group1| (N=&N1)" right;
define _2 / display "&group2| (N=&N2)" right;

```

```

define pvalue / display "p-value" right;

compute sv;
if sv=0 then
call define(_row_, 'style', 'style=[font_weight=bold]');
endcomp;

title "&table. Green Dot Intervention and Demographics of Snapshot in &group1
and &group2";
footnotel j=center height=9pt "Biostatistics Capstone--Life's Snapshot
Project. &SYSDATE9";
quit;
%mend;

ods rtf file="M:\MPH\2018 Fall Semester\CPH 608 Capstone\notes\reports.rtf";
proc report data=FRQ headline headskip nowindows split='|';
columns fupgp col CumFrequency CumPercent;
define fupgp / display "Follow-up Completed Groups";
define col / display "N(%)";
define CumFrequency / display "Cumulative Frequency";
define CumPercent / display "Cumulative Percentage";
title1 "Table 1. Descriptive Estimates Rates";
footnote j=center height=9pt "Biostatistics Capstone--Life's Snapshot
Project. &SYSDATE9";
run;
%report(final1,Baseline Only,At Least One Follow-up,&GP1,&GP2,Table 2.);
%report(final2,Baseline Only,Five Follow-ups,&GP1,&GP3,Table 3.);
%report(final3,Baseline Only,Final Finished,&GP1,&GP4,Table 4.);
%report(final4,At Least One Follow-up in Four,Only Final,&GP5,&GP6,Table 5.);
ods rtf close;

ods rtf file="M:\MPH\2018 Fall Semester\CPH 608
Capstone\notes\logistic_reports.rtf";
proc logistic data=merged;
class GDMATYN gender RUCCGP NONWHITE POVGP apyn;
model surind1=GDMATYN gender RUCCGP NONWHITE POVGP apyn;
format
surind1 surind_a.
GDMATYN gdmatyn.
gender gender.
RUCCGP resarea.
NONWHITE nonwhite.
POVGP poverty.
apyn apyn.;
run;

proc logistic data=merged;
class GDMATYN gender NONWHITE apyn / param=ref;
model surind1=GDMATYN gender NONWHITE apyn;
format
surind1 surind_a.
GDMATYN gdmatyn.
gender gender.
NONWHITE nonwhite.
apyn apyn.;
run;
ods rtf close;

```

Project Outcomes (Tables and Graphs)

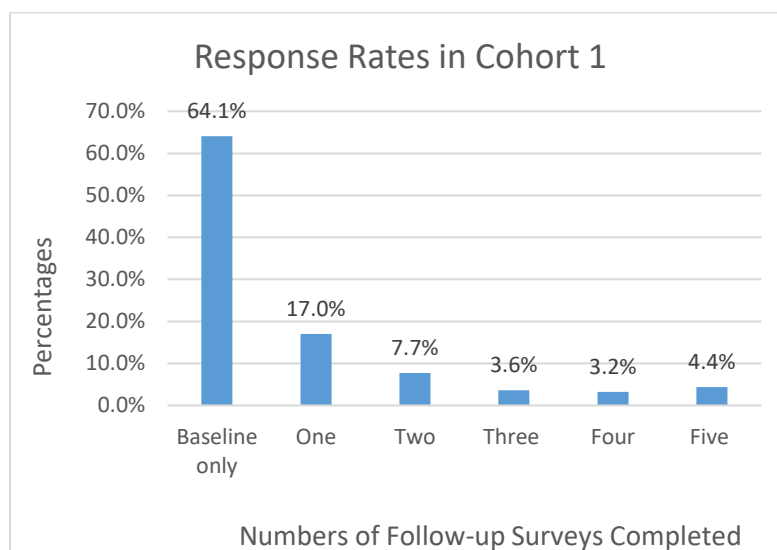
Descriptive Tables and Graphs

Table 1. Descriptive Estimates Rates

Follow-up Completed Groups	N (%)	Cumulative Frequency	Cumulative Percentage
Baseline only	2260 (64.1%)	2260	64.08
One follow-up completed	601 (17.0%)	2861	81.12
Two follow-ups completed	271 (7.7%)	3132	88.80
Three follow-ups completed	127 (3.6%)	3259	92.40
Four follow-ups completed	113 (3.2%)	3372	95.61
Five follow-ups completed	155 (4.4%)	3527	100.00

Table 1 provides the frequency and percentage of different numbers of follow-ups completed. According to this table, after the baseline survey, 64.1% of participants were lost to follow up. Hence, a comparison between these people and people who completed at least one follow-up might provide some information on retention. However, there were more than 100 participants finished all the follow-up surveys. The demographics of this group of people would be a reference for high retention rates. In addition, people who finished the final survey were also compared to other groups of people to explore the motivation of incentive. Figure 1 shows that the response rates decrease as the number of follow-up finished increases. However, a slight increase at the end indicates that the incentive works to increase the response.

Figure 1. Descriptive Estimates Rates

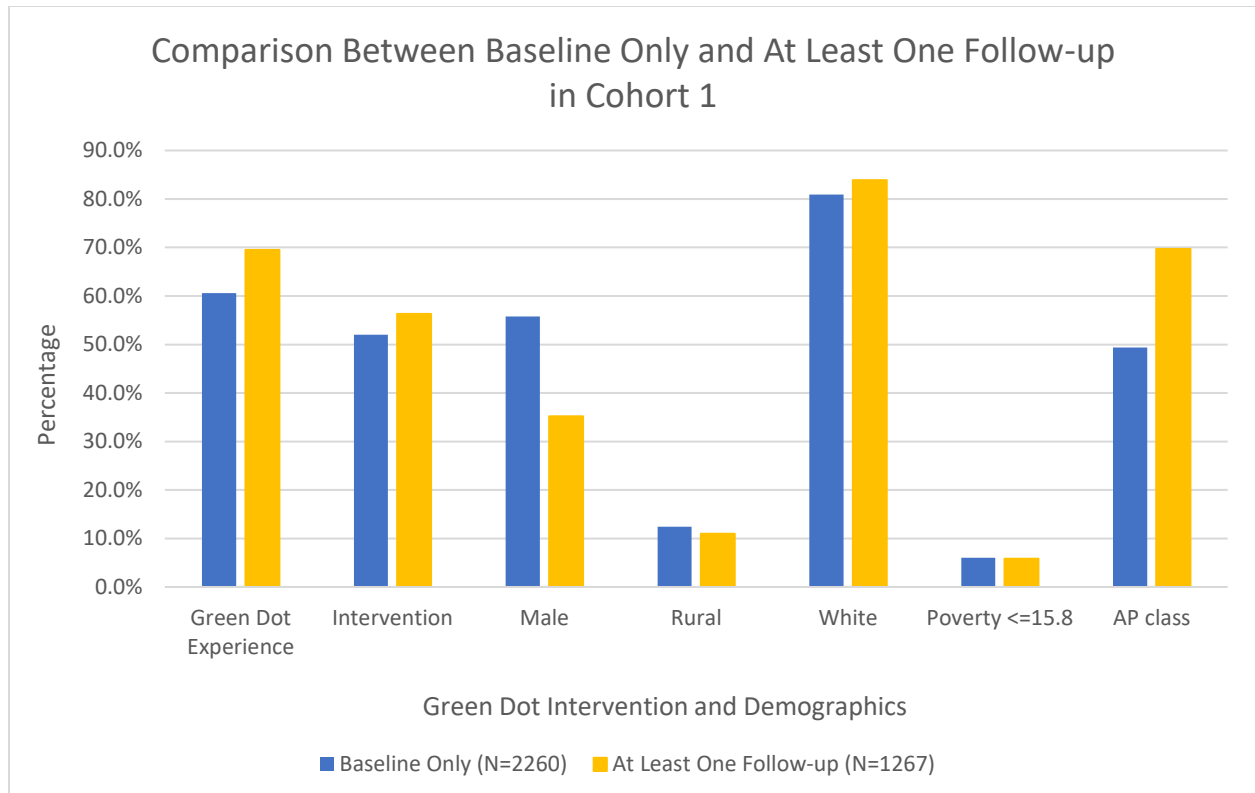


**Table 2. Green Dot Intervention and Demographics of Snapshot in
Baseline Only and At Least One Follow-up**

	Baseline Only (N=2260)	At Least One Follow-up (N=1267)	p-value
GDMATYN			<.0001
No Experience	890 (39.4%)	386 (30.5%)	
Green Dot Training, Speeches or Other	1370 (60.6%)	881 (69.5%)	
INT			0.0144
Control	1085 (48.0%)	554 (43.7%)	
Intervention	1175 (52.0%)	713 (56.3%)	
GENDER			<.0001
Female	998 (44.2%)	817 (64.8%)	
Male	1259 (55.8%)	443 (35.2%)	
RUCCGP			0.3549
Urban	1116 (49.4%)	618 (48.8%)	
Suburban	864 (38.2%)	509 (40.2%)	
Rural	280 (12.4%)	140 (11.0%)	
NONWHITE			0.0226
White	1816 (80.9%)	1056 (83.9%)	
Non-white	430 (19.1%)	202 (16.1%)	
POVGP			0.9641
<=15.8	136 (6.0%)	74 (5.8%)	
15.8 - 17.8	201 (8.9%)	115 (9.1%)	
>17.8	1923 (85.1%)	1078 (85.1%)	
APYN			<.0001
No	1138 (50.6%)	382 (30.3%)	
Yes	1109 (49.4%)	880 (69.7%)	

Table 2 presents the results of the Chi-square test between baseline only group and at least one follow-up completed group. Green Dot experience is significantly different in two follow up groups, indicating people who know about the Green Dot Program tend to have a higher rate of response. Gender and race are significantly different in the two groups as well while other demographics does not influence the follow-up retention. The p-value for AP class indicates that students who were planning to go to college were more likely to respond to follow-up surveys.

**Figure 2. Green Dot Intervention and Demographics of Snapshot in
Baseline Only and At Least One Follow-up**

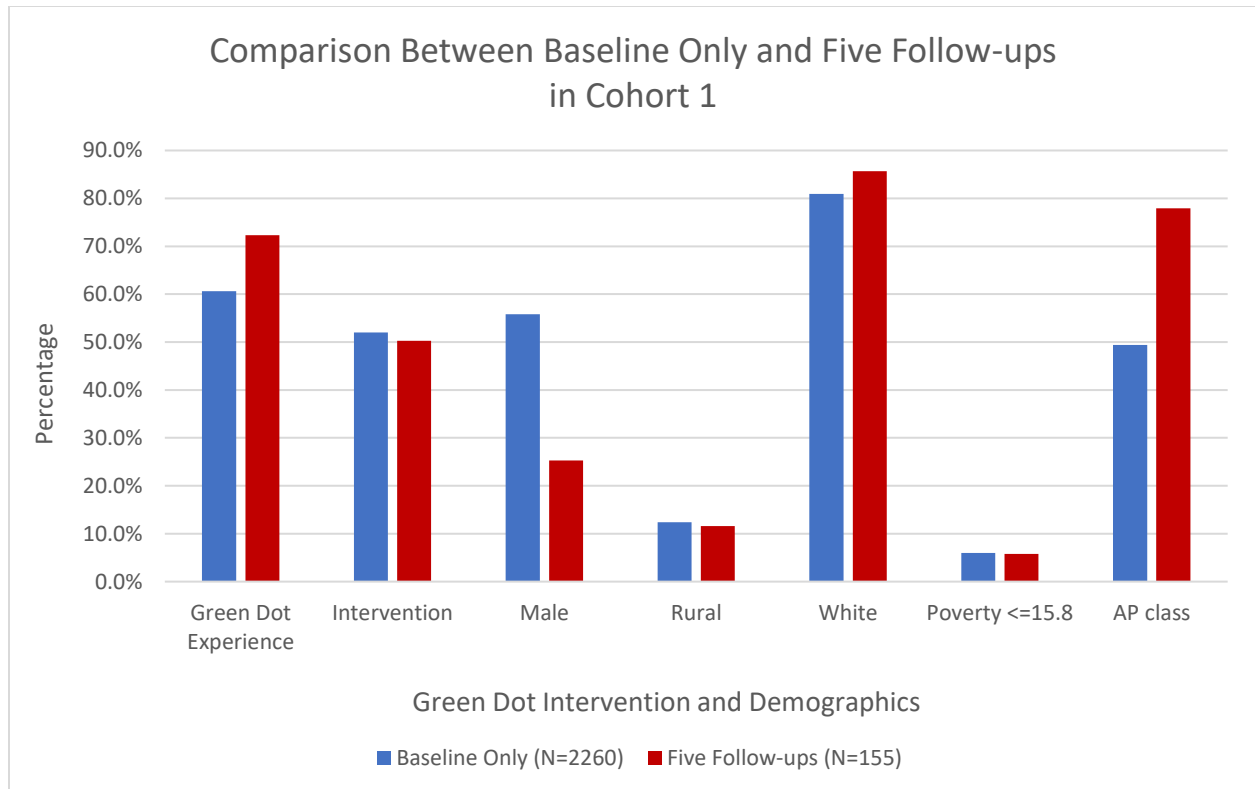


**Table 3. Green Dot Intervention and Demographics of Snapshot in
Baseline Only and Five Follow-ups**

	Baseline Only (N=2260)	Five Follow-ups (N=155)
GDMATYN		
No Experience	890 (39.4%)	43 (27.7%)
Green Dot Training, Speeches or Other	1370 (60.6%)	112 (72.3%)
INT		
Control	1085 (48.0%)	77 (49.7%)
Intervention	1175 (52.0%)	78 (50.3%)
GENDER		
Female	998 (44.2%)	115 (74.7%)
Male	1259 (55.8%)	39 (25.3%)
RUCCGP		
Urban	1116 (49.4%)	71 (45.8%)
Suburban	864 (38.2%)	66 (42.6%)
Rural	280 (12.4%)	18 (11.6%)
NONWHITE		
White	1816 (80.9%)	132 (85.7%)
Non-white	430 (19.1%)	22 (14.3%)
POVGP		
<=15.8	136 (6.0%)	9 (5.8%)
15.8 - 17.8	201 (8.9%)	16 (10.3%)
>17.8	1923 (85.1%)	130 (83.9%)
APYN		
No	1138 (50.6%)	34 (22.1%)
Yes	1109 (49.4%)	120 (77.9%)

Notes: P-values were not calculated due to small number of completing five follow-ups. Results are provided as descriptive only.

**Figure 3. Green Dot Intervention and Demographics of Snapshot in
Baseline Only and Five Follow-ups**



**Table 4. Green Dot Intervention and Demographics of Snapshot in
Baseline Only and Final Finished**

	Baseline Only (N=2260)	Final Finished (N=432)
GDMATYN		
No Experience	890 (39.4%)	131 (30.3%)
Green Dot Training, Speeches or Other	1370 (60.6%)	301 (69.7%)
INT		
Control	1085 (48.0%)	198 (45.8%)
Intervention	1175 (52.0%)	234 (54.2%)
GENDER		
Female	998 (44.2%)	293 (68.1%)
Male	1259 (55.8%)	137 (31.9%)
RUCCGP		
Urban	1116 (49.4%)	215 (49.8%)
Suburban	864 (38.2%)	168 (38.9%)
Rural	280 (12.4%)	49 (11.3%)
NONWHITE		
White	1816 (80.9%)	363 (84.4%)
Non-white	430 (19.1%)	67 (15.6%)
POVGP		
<=15.8	136 (6.0%)	28 (6.5%)
15.8 - 17.8	201 (8.9%)	31 (7.2%)
>17.8	1923 (85.1%)	373 (86.3%)
APYN		
No	1138 (50.6%)	118 (27.4%)
Yes	1109 (49.4%)	313 (72.6%)
Number of Follow-ups completed		
0	N/A	32 (7.42%)
1-2	N/A	150 (34.72%)
3-4	N/A	250 (57.87%)

Notes: P-values were not calculated due to small number of participants in some categories. Results are provided as descriptive only.

**Table 5. Green Dot Intervention and Demographics of Snapshot in
At Least One Follow-up in Four and Only Final**

	At Least One Follow-up in Four (N=1235)	Only Final (N=32)
GDMATYN		
No Experience	376 (30.4%)	10 (31.3%)
Green Dot Training, Speeches or Other	859 (69.6%)	22 (68.8%)
INT		
Control	540 (43.7%)	14 (43.8%)
Intervention	695 (56.3%)	18 (56.3%)
GENDER		
Female	802 (65.3%)	15 (46.9%)
Male	426 (34.7%)	17 (53.1%)
RUCCGP		
Urban	600 (48.6%)	18 (56.3%)
Suburban	498 (40.3%)	11 (34.4%)
Rural	137 (11.1%)	3 (9.4%)
NONWHITE		
White	1029 (83.9%)	27 (84.4%)
Non-white	197 (16.1%)	5 (15.6%)
POVGP		
<=15.8	70 (5.7%)	4 (12.5%)
15.8 - 17.8	113 (9.1%)	2 (6.3%)
>17.8	1052 (85.2%)	26 (81.3%)
APYN		
No	367 (29.8%)	15 (46.9%)
Yes	863 (70.2%)	17 (53.1%)

Notes: P-values were not calculated due to small number of completing final only. Results are provided as descriptive only.

Table 5 presents descriptive statistics for people who finished the first four follow-ups and people who only completed the final survey, which has \$50 as incentive. However, very small group of people finished only the final survey.

Table 6. Comparison Between Follow-up Groups

	Baseline Only (N=2260)	At Least One Follow-up (N=1267)	p-value	Five Follow-ups (N=155)
GDMATYN			<.0001	
No Experience	890 (39.4%)	386 (30.5%)		43 (27.7%)
Green Dot Training, Speeches or Other	1370 (60.6%)	881 (69.5%)		112 (72.3%)
INT			0.0144	
Control	1085 (48.0%)	554 (43.7%)		77 (49.7%)
Intervention	1175 (52.0%)	713 (56.3%)		78 (50.3%)
GENDER			<.0001	
Female	998 (44.2%)	817 (64.8%)		115 (74.7%)
Male	1259 (55.8%)	443 (35.2%)		39 (25.3%)
RUCCGP			0.3549	
Urban	1116 (49.4%)	618 (48.8%)		71 (45.8%)
Suburban	864 (38.2%)	509 (40.2%)		66 (42.6%)
Rural	280 (12.4%)	140 (11.0%)		18 (11.6%)
NONWHITE			0.0226	
White	1816 (80.9%)	1056 (83.9%)		132 (85.7%)
Non-white	430 (19.1%)	202 (16.1%)		22 (14.3%)
POVGP			0.9641	
<=15.8	136 (6.0%)	74 (5.8%)		9 (5.8%)
15.8 - 17.8	201 (8.9%)	115 (9.1%)		16 (10.3%)
>17.8	1923 (85.1%)	1078 (85.1%)		130 (83.9%)
APYN			<.0001	
No	1138 (50.6%)	382 (30.3%)		34 (22.1%)
Yes	1109 (49.4%)	880 (69.7%)		120 (77.9%)

*Groups are not mutually exclusive

*P-values only computed for comparison of baseline only to at least one follow-up.

Logistic Regression

The LOGISTIC Procedure

Model Information	
Data Set	WORK.MERGED
Response Variable	surind1
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	3527
Number of Observations Used	3487

Response Profile		
Ordered Value	surind1	Total Frequency
1	At Least One Follow-up	1253
2	Baseline Only	2234

Probability modeled is surind1='At Least One Follow-up'.

Note: 40 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information			
Class	Value	Design Variables	
GDMATYN	Green Dot Training, Speeches or Other	1	
	No Experience	-1	
gender	Female	1	
	Male	-1	
RUCCGP	Rural	1	0
	Suburban	0	1
	Urban	-1	-1
NONWHITE	Non-white	1	
	White	-1	

Class Level Information			
Class	Value	Design Variables	
POVGP	15.8 - 17.8	1	0
	<=15.8	0	1
	>17.8	-1	-1
apyn	No	1	
	Yes	-1	

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4556.262	4293.205
SC	4562.419	4348.616
-2 Log L	4554.262	4275.205

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	279.0568	8	<.0001
Score	268.5469	8	<.0001
Wald	252.8243	8	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
GDMATYN	1	13.9725	0.0002
gender	1	112.8439	<.0001
RUCCGP	2	0.4434	0.8012
NONWHITE	1	6.4354	0.0112

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
POVGP	2	1.2063	0.5471
apyn	1	104.3900	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8636	0.0925	87.1128	<.0001
GDMATYN	Green Dot Training, Speeches or Other	1	0.1485	0.0397	13.9725	0.0002
gender	Female	1	0.3956	0.0372	112.8439	<.0001
RUCCGP	Rural	1	-0.0298	0.0792	0.1416	0.7067
RUCCGP	Suburban	1	0.0384	0.0583	0.4353	0.5094
NONWHITE	Non-white	1	-0.1329	0.0524	6.4354	0.0112
POVGP	15.8 - 17.8	1	-0.0998	0.1003	0.9891	0.3200
POVGP	<=15.8	1	0.0555	0.1152	0.2318	0.6302
apyn	No	1	-0.3965	0.0388	104.3900	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
GDMATYN Green Dot Training, Speeches or Other vs No Experience	1.346	1.152	1.572
gender Female vs Male	2.206	1.906	2.553
RUCCGP Rural vs Urban	0.979	0.764	1.255
RUCCGP Suburban vs Urban	1.048	0.885	1.242
NONWHITE Non-white vs White	0.767	0.624	0.941
POVGP 15.8 - 17.8 vs >17.8	0.866	0.668	1.122
POVGP <=15.8 vs >17.8	1.011	0.732	1.397
apyn No vs Yes	0.452	0.389	0.527

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	64.5	Somers' D	0.328
Percent Discordant	31.7	Gamma	0.341
Percent Tied	3.8	Tau-a	0.151
Pairs	2799202	c	0.664

The probability of at least one follow-up completed was modeled in the first logistic regression with GDMATYN, intervention and demographics. Residential area and poverty are unrelated to the follow-up retention. Hence, they were excluded in the second logistic regression.

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4556.262	4286.654
SC	4562.419	4317.438
-2 Log L	4554.262	4276.654

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	277.6082	4	<.0001
Score	267.3131	4	<.0001
Wald	251.8053	4	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
GDMATYN	1	13.0393	0.0003
gender	1	112.4301	<.0001
NONWHITE	1	7.6107	0.0058
apyn	1	108.5383	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8263	0.0840	96.8644	<.0001
GDMATYN	Green Dot Training, Speeches or Other	1	0.2827	0.0783	13.0393	0.0003
gender	Female	1	0.7888	0.0744	112.4301	<.0001
NONWHITE	Non-white	1	-0.2696	0.0977	7.6107	0.0058
apyn	No	1	-0.7959	0.0764	108.5383	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
GDMATYN Green Dot Training, Speeches or Other vs No Experience	1.327	1.138	1.547
gender Female vs Male	2.201	1.902	2.546
NONWHITE Non-white vs White	0.764	0.631	0.925
apyn No vs Yes	0.451	0.388	0.524

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	61.4	Somers' D	0.325
Percent Discordant	28.9	Gamma	0.360
Percent Tied	9.6	Tau-a	0.150
Pairs	2799202	c	0.663

According to the output of this logistic regression, Green Dot experience was associated with increased rate of retention (odds ratio (OR) = 1.327, 95% confidence interval (CI): 1.138, 1.547). Females are related to higher rate of retention (odds ratio (OR) = 2.201, 95% confidence interval (CI): 1.902, 2.546) while non-whites are associated with reduced rate of retention (odds ratio (OR) = 0.764, 95% confidence interval (CI): 0.631, 0.925). In addition, students who took AP class are more likely to response than those who did not take AP class.

Lessons Learned

I have been learning, applying, and extending theoretical knowledge obtained from the MPH program throughout the capstone project. As a Biostatistics concentrator, I will summarize the lessons learned in the following aspects.

Study Design

Longitudinal studies follow the same samples at several time points to capture change at the individual level, which help find patterns that may occur over long periods and discover relationships between exposure and outcome. However, by working with real-world data, I learned that a big problem in this type of study is the loss to follow-up. This problem can cause bias and reduce study power, influencing results. Some strategies would be helpful to improve retention, including questionnaire format, communication strategies, incentives, and case management strategies. Sampling is also an important part study design.

Statistical Methodologies

I reinforced descriptive statistics and inferential statistics in the data analysis phase. To explore the relationship between two categorical variables, we can use the Chi-square test. What should we be aware of is information extraction from results according to our purposes. The crosstabulation table includes frequencies and percentages. We should be able to extract right information based on the dependent variable and independent variable. In addition, I have a better understanding of logistic regression. I used logistic regression usually for prediction. However, it can be used to identify significant factors without adjusting confounders.

Statistical Analysis

I made significant progress through this project on advanced SAS programming. Since I had multiple repeated procedures, I learned how to use Macro in the analysis to obtain results in a more effective approach. By exploring Macro, I acquired various ways to define macro variables, including “%global statement”, “%let statement”, SQL “select into:”, and macro parameters. In addition, it was the first time that I knew SAS procedures assign names to each table they generate. We can use those names to reference tables when using the Output Delivery System (ODS) to select tables and create output data sets. In logistic regression, I learned about dummy variables, “descending” option, “(ref=’)” option, “param= ” option to control the response level or reference level.

Results

The results of the statistical tests are not proper to be directly listed for reports. Using Macro to generate the summary report would be the most effective way, avoiding manual errors. By combining macro and ODS, I am able to generate report tables by SAS. Titles, labels, footnotes were also added to reports to make them more readable and understandable. Output Delivery System (ODS) Destinations include RTF, PDF, HTML, EXCEL, etc, which can help generate different types of reports.

Project Summary

Long-term follow-up is necessary to evaluate the effect of violence prevention programs. However, the loss to follow-up is one of the common problems in prospective longitudinal studies. This project aims to investigate factors affecting retention of follow-up in Life's Snapshot Project.

Chi-square tests and logistic regression were performed. The results show that Green Dot experience, AP class, gender, and race are significantly associated with rates of response. People who have Green Dot experience and who take AP class are related to higher rates of retention. Females are associated with higher retention rates while non-whites are associated with lower rates.

This project can provide a reference for the further analysis in Life's Snapshot Project. Since the population in the follow-up are not representative, we should take this into account in the analysis of the effectiveness of Green Dot intervention. What is more, this project would provide strategies for representative sampling and retention improvement in future longitudinal studies.